

Structure of local interactions in complex financial dynamics

X.F. Jiang, T.T. Chen, and B. Zheng*

Department of Physics, Zhejiang University, Hangzhou 310027, P.R. China

(Dated: June 3, 2014)

With the network methods and random matrix theory, we investigate the interaction structure of communities in financial markets. In particular, based on the random matrix decomposition, we clarify that the local interactions between the business sectors (subsectors) are mainly contained in the sector mode. In the sector mode, the average correlation inside the sectors is positive, while that between the sectors is negative. Further, we explore the time evolution of the interaction structure of the business sectors, and observe that the local interaction structure changes dramatically during a financial bubble or crisis.

arXiv:1406.0070v1 [q-fin.GN] 31 May 2014

* zheng@zimp.zju.edu.cn

INTRODUCTION

Financial markets are dynamic systems with complex interactions, which share common features in various aspects with those in traditional physics. In recent years, large amounts of historical data have piled up in stock markets. This allows a quantitative analysis of the financial dynamics with the concepts and methods in statistical mechanics, and many empirical results have been documented [1–24]. Very recently, these results are also supported and augmented with the online big data. For examples, the price change may be predicted with the mood on Twitter [25], and the trading behavior may be quantified with the Google search volume and Wikipedia topic view times [26, 27]. Statistical properties of the price fluctuations and cross-correlations between individual stocks attract much attention, not only scientifically for unveiling the complex structure and internal interactions of the financial system, but also practically for the asset allocation and portfolio risk estimation [5, 28].

In this paper, we are mainly concerned with the cross-correlations between individual stocks in financial markets. Previous researches have revealed that the cross-correlation between stocks fluctuates over time [29–31]. In particular, it could be enhanced during the recession of business cycles. Similarly, the international correlation is stronger in a volatile period. Based on the time-lag cross-correlations, one obtains that changes in singular eigenvectors and eigenvalues are the largest during a financial crisis [13]. Further, individual stocks may form communities due to the cross-correlations between stocks. A community is also called a business sector, since its stocks usually share common economic properties. For example, the business sectors have been investigated with the random matrix theory (RMT), in the mature markets such as the American stock market and Korean stock market [32–37], and also in the emerging markets such as the Chinese stock market and Indian stock market [10, 16, 38]. In the emerging stock markets, both standard and unusual business sectors are detected. The RMT theory may identify the business sectors in a financial market, but it could not draw the interactions between the business sectors. Especially, the global price movement in a stock market is usually much stronger than the local one. To uncover the local interactions between the business sectors, therefore, it is necessary to develop a proper method to subtract the global price movement. On the other hand, the structure of financial networks has been gaining increasing interests [39–41]. For example, the correlation structure of individual stocks has been investigated with the minimal spanning tree (MST) and planar maximally filtered graph (PMFG) [17, 42–46].

The main motivation of this paper is to investigate the interaction structure between the business sectors in complex financial systems. Especially, we extract the local interactions between the business sectors with the random matrix decomposition. The methodology is to combine the RMT theory with various methods in complex networks, such as the MST tree and PMFG graph, and the theoretical information method (Infomap). In a complex network, the community structure is the gathering of nodes into communities with a very high density of internal edges, including a relatively high density of edges between related communities. Exploring the community structure with the concepts and methods in statistical mechanics is a very important topic in the field of complex networks [47–50]. Our study in this paper will provide a deeper understanding on the communities, i.e., the business sectors, the interactions between the business sectors, and how this business-sector structure emerges from the correlations between individual stocks. In particular, with the random matrix decomposition based on the RMT theory, we clarify that the local interactions between the business sectors are mainly contained in the sector mode. In the sector mode, the average correlation inside the sectors is positive, while that between the sectors is negative. Further, we study the time evolution of the interaction structure. We discover that the interaction structure of the sector mode varies dramatically with time, while that of the full cross-correlation matrix does not. Particularly, the interaction structure of the business sectors in the sector mode is simple during a financial bubble or crisis, and it is dominated by specific business sectors, which are associated with the corresponding economic situation.

RESULTS

Interaction structure of business sectors

Recently a number of activities have been devoted to the study of the statistical and topological properties of the financial networks [17, 42–46]. In this paper we concentrate our attention on the interactions between the communities. Two financial networks, i.e., the MST tree and PMFG graph are first generated from the cross-correlation matrix of the financial market. Then the Infomap method is applied to capture the interaction structure of the communities from the MST tree and PMFG graph. In common terminology, a community in a financial market is also called a business sector, since its stocks usually share common economic properties. In the RMT theory, a business sector may split into the positive and negative subsectors, which are anti-correlated each other [16]. In the present paper, such a phenomenon is also observed in the community structure of the sector mode obtained from the MST tree and PMFG graph. In these cases, a subsector is usually considered to be a single community, but for convenience, we may

still use the word “a business sector” to represent “a business sector (subsector)”.

To investigate the cross-correlations between individual stocks and the community structure in a stock market, the data set should contain as many stocks as we can obtain. On the other hand, the total time length of the time series of stock prices should also be as long as possible. With these considerations, we obtain the daily data of 259 weighted stocks in the Shanghai Stock Exchange (SSE), from Jan., 1997 to Nov., 2007, with 2633 data points in total. For comparison, we select also 259 weighted stocks in the New York Stock Exchange (NYSE), and collect the daily data from Jan., 1990 to Dec., 2006, with 4286 data points in total. In the selection of these 259 stocks, we keep the diversity cross different business sectors. In addition, to further analyze the interaction structure of the communities during a financial crisis, we collect another data set, with the daily data of 249 weighted stocks from Oct., 2007 to Nov., 2008, for both the NYSE and SSE markets. Over ninety-five percent of these 249 stocks overlap with the above 259 stocks. All these data are obtained from Yahoo Finance (<http://finance.yahoo.com>). If the price of a stock in a particular day is absent, it is assumed that the price is the same as the preceding day [56]. It has been pointed out that the missing data do not lead to artifacts [38].

In Fig. 1, the community structure of the PMFG graph for the NYSE market is shown. That of the MST tree looks similar but without cycles. Most of the communities of both the MST tree and PMFG graph can be identified as the business sectors detected with the RMT theory [10, 16]. In the NYSE market, the business sectors are usually standard. In the SSE market, however, there exist a few unusual business sectors, whose effects are even stronger than the standard ones [10, 16]. For example, a company will be specially treated if its financial situation is abnormal. In this case, the acronym “ST” will then be added to the stock name as a prefix. The abnormal financial situation includes: the audited profits are negative in two successive accounting years, and the audited net worth per share is less than the par value of the stock in the recent accounting year etc. The acronym “ST” will be removed when the financial situation becomes normal. The so-called ST sector consists of the “ST” stocks [10].

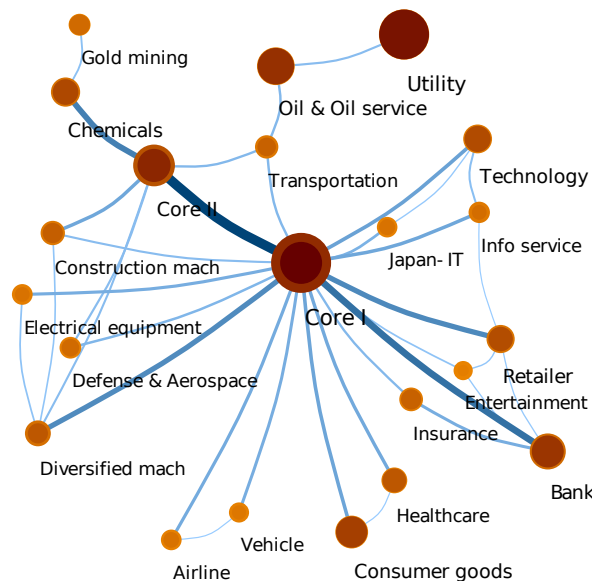


FIG. 1. The interaction structure of the business sectors of the PMFG graph for the NYSE market, plotted with an online tool of MapEquation (<http://www.mapequation.org>). The size of a node represents the number of stocks inside a community, and the width of an edge indicates the strength of the interaction between two communities.

In Table I, the business sectors of the PMFG graph are compared with those from the RMT theory. In fact, all business sectors (subsectors) identified with the RMT theory can be observed in the communities of the PMFG graph, but it is not true vice versa. A typical example in the NYSE market is, that four communities of industrial goods, including Diversity machinery, Construction machinery, Defense & Aerospace, and Electrical equipment, can not be found in the RMT sectors. On the other hand, the stocks in different RMT sectors may overlap each other, but this is not allowed in the communities of the PMFG graph. In the RMT theory, for examples, both λ_6^- and λ_9^+ represent the finance business sectors, which share most common stocks.

To quantitatively characterize the interaction structure of the business sectors, we define the average correlation inside a business sector as

$$\bar{C}_{ij}^{in} = \frac{1}{E} \sum_p \sum_{i \neq j} C_{ij}(p), \quad (1)$$

where $C_{ij}(p)$ is the cross-correlation C_{ij} of an edge between i and j stocks belonging to a same sector p , and E is the total number of such edges. For comparison, we define the average correlation between two business sectors as

$$\bar{C}_{ij}^{be} = \frac{1}{E'} \sum_{p \neq q} \sum_{i \neq j} C_{ij}(pq), \quad (2)$$

where $C_{ij}(pq)$ is the cross-correlation C_{ij} of an edge between i and j stocks belonging to different business sectors p and q , and E' is the total number of such edges. The results of \bar{C}_{ij}^{in} and \bar{C}_{ij}^{be} are shown in Table II. For the NYSE market, \bar{C}_{ij}^{in} of the RMT sectors (subsectors), MST tree and PMFG graph, are 0.33, 0.49 and 0.40 respectively, which are two to three times as large as the average cross-correlation, $\bar{C}_{ij} = 0.16$. For the SSE market, \bar{C}_{ij}^{in} are 0.40, 0.48 and 0.43 respectively, which should be compared with the average cross-correlation, $\bar{C}_{ij} = 0.37$. We also note that \bar{C}_{ij}^{in} of the MST tree and PMFG graph are larger than those of the RMT sectors. In this sense, the network methods refine the business sectors obtained with the RMT theory. By the definition of the communities, the average correlation inside a community, \bar{C}_{ij}^{in} , should be stronger than that between two communities, \bar{C}_{ij}^{be} . Since there are ten to twenty communities, the total number of edges between two different communities is much larger than the total number of edges inside a community. Therefore, as can be seen in Table II, \bar{C}_{ij}^{be} is very close to \bar{C}_{ij} , much smaller than \bar{C}_{ij}^{in} .

To further understand the interactions between the business sectors in Fig. 1, we introduce \bar{C}_{ij}^i to denote the average correlation of the edges between two business sectors with a link, and \bar{C}_{ij}^{de} to represent the one without links. For comparison, the average correlation between the positive and negative subsectors in a particular eigenmode of the RMT theory, \bar{C}_{ij}^{+-} , is also computed[16]. All these results are shown in Table II. It is clear that the correlation \bar{C}_{ij}^{de} is weaker than \bar{C}_{ij}^i . In fact, \bar{C}_{ij}^{be} is a weighted average of \bar{C}_{ij}^i and \bar{C}_{ij}^{de} . For the RMT theory, \bar{C}_{ij}^{+-} is not only much smaller than \bar{C}_{ij}^{in} , but also somewhat smaller than \bar{C}_{ij}^{be} , although both the positive and negative subsectors formally belong to a same sector. This is a direct indication of the anti-correlation between the positive and negative subsectors in a particular eigenmode [16].

To the best of our knowledge, this is the first time that one reveals the interaction structure of the business sectors in the financial markets. In a certain sense, the community structure of the MST tree or PMFG graph may provide a more comprehensive understanding of the interactions between the business sectors, compared to the RMT theory. However, we observe that in the community structure of the PMFG graph in Fig. 1, there are two dominating communities, i.e., the so-called Core I and Core II, which can not be identified as the standard business sectors. Careful examination shows that the stocks in Core I and Core II are mainly those with relatively large components in the eigenvector of the largest eigenvalue in the RMT theory. In fact, those stocks may be identified as the controlling set of the financial network with a spectral centrality method, and Core I and Core II are the network drivers, which control the pace of the entire system [51–55]. To some extent, Core I and Core II form a kind of centers in the community structure, connecting to the majority of the business sectors. In contrast, there are only a few links between other standard business sectors. In other words, the interaction structure of the business sectors in Fig. 1 is dominated by Core I and Core II, but it does not represent the real local interactions between the business sectors. Therefore, it is necessary to exclude the global price movement, and to extract the real local interactions between the business sectors.

Local interaction structure in sector mode

According to the random matrix decomposition described by Eq. (7) in Sec. Methods, we may decompose the cross-correlation matrix into N eigenmodes. With Eq. (8), these eigenmodes are then grouped into three important modes, which are called the market mode, sector mode and random mode in this paper. In Fig. 2, the probability distributions of the matrix elements of the three modes are shown. The probability distribution of the matrix elements of the market mode is similar to that of the full cross-correlation matrix, while the central peaks of the sector mode and random mode are close to zero.

Now we define a domain as a collection of stocks, in which all cross-correlations between the stocks are with a same sign, i.e., positive or negative. The domain formed by positive or negative correlations is called the positive or negative domain, respectively. For the sector mode, the probability distributions of the domain size are shown in Fig. 3. For the NYSE market, there are 17 positive domains, the largest domain size is 48, and the average size is about 15. While there are 61 negative domains, the largest domain size is 7, and the average size is about 4. The probability distribution of the domain size of the SSE market is similar to that of the NYSE market. If a particular stock connects to other two stocks with negative correlations, there is a low possibility that the correlation between

those two domains is also negative. Thus, it is natural that the average size of the negative domains is small. The probability distribution of the domain size of both the positive and negative domains for the random mode is similar to that of the negative domains for the sector mode. These results imply that the community structure should be mainly contained in the positive domains of the sector mode.

In fact, the positive domains of the sector mode can be associated with the business sectors, which are almost identical to those detected with the RMT theory shown in Table I. On the other hand, one can not extract the business sectors from the negative domains of the sector mode, as well as both the positive and negative domains of the random mode. For the market mode or the full cross-correlation matrix, there are very few domains since almost all the correlations are positive, and these domains are also irrelevant to the business sectors.

To clearly show the domains of the sector mode, and to investigate their interactions, we plot the interactions between the individual stocks in Fig. 4. Gray and white dots denote positive and negative correlations between two stocks, respectively. The squares along the diagonal line are the domains. To uncover the interaction structure of the domains, we calculate the average correlation $\overline{C}_{sec}^{be}(pq)$ between two different domains p and q . In Fig. 4, two domains will be linked with a line, if $\overline{C}_{sec}^{be}(pq)$ between this pair of domains is larger than the average value of $\overline{C}_{sec}^{be}(pq)$ for all different pairs of domains.

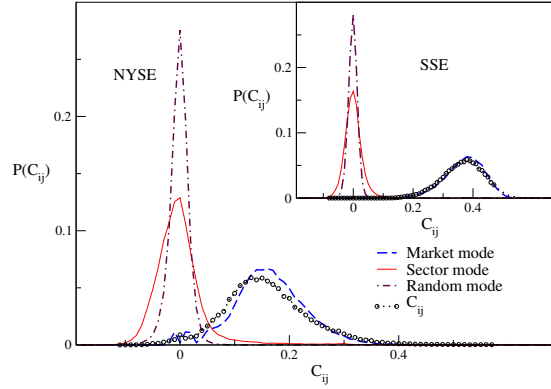


FIG. 2. Probability distributions of the matrix elements C_{ij} and C_{ij}^{mode} for the NYSE and SSE markets.

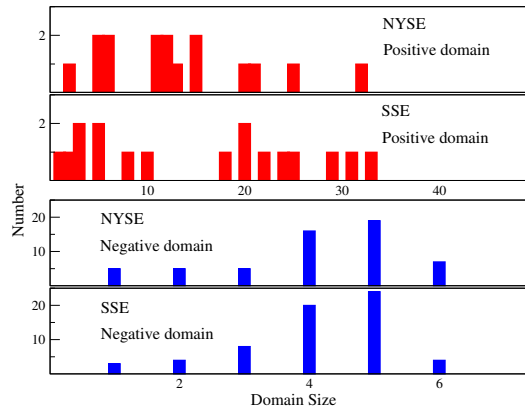


FIG. 3. Probability distributions of the domain size in the sector mode for the NYSE and SSE markets.

To further understand the business sectors and their local interactions in the sector mode, we follow the procedure in the previous section, to generate the PMFG graph from the absolute values of the matrix elements of the sector-mode

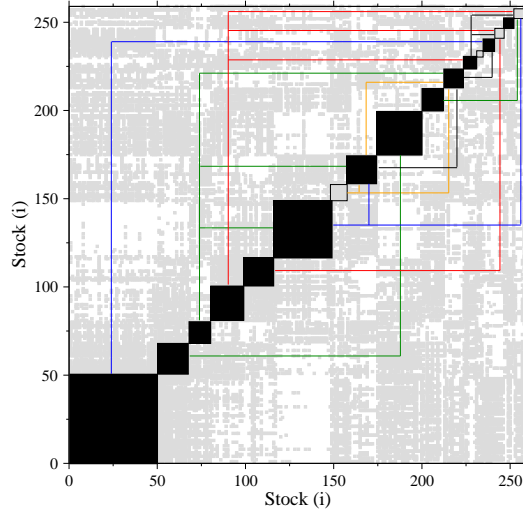


FIG. 4. The domain structure for the NYSE market. Gray and white dots denote positive and negative correlations between two stocks, respectively. The squares along the diagonal line are the domains. The solid black squares represent the domains which could be associated with the business sectors, while others could not be identified as the standard business sectors. The business sectors from left corner to right are: Industrial goods, Consumer goods, Oil I, Retailer, Finance, Utility, Oil II, IT, Food & Personal care, Metal Mining, Healthcare, Machinery and Transportation. Here Oil I and Oil II are two groups of Oil companies.

cross-correlation matrix C_{sec} defined in Eq. (8), and to extract the community structure from the PMFG graph with the Infomap method. The communities of the PMFG graph of the sector mode can also be identified as the business sectors, which are almost the same as those from the full cross-correlation matrix C . As shown in Fig. 5, however, the interaction structure of the business sectors of the sector mode is very different from that of the full cross-correlation matrix. Core I and Core II in Fig. 1 do not show up anymore. In contrast, IT, IG, and Food & Personal care sectors occupy the central positions for the NYSE market, and IT-BM and Utility-Finance sectors are the hubs for the SSE market. In other words, these business sectors actually play a key role in the interaction structures of the business sectors of the sector mode. Here, we observe that the interaction structure between the business sectors in Fig. 5 is similar to that in Fig. 4.

To briefly summarize, with the random matrix decomposition, one may remove the global price movement and background noisy correlations described by the market mode and random mode respectively, and identify the local interaction structure of the business sectors in the sector mode. The sector mode which reflects the local interactions of the business sectors, plays an important role in the financial market. On the other hand, we should note that one can not obtain meaningful business sectors from the market mode and random mode.

Additionally, looking carefully at the internal interactions in a community, one may find that a community may split into two subcommunities, which are connected by the negative correlations. These two subcommunities are called a cluster pair, which is somewhat similar to the pair of positive and negative subsector in the RMT theory [16]. In Fig. 5, two semi-circles with different colors adhered in a full circle represent a cluster pair. Four cluster pairs are identified for the NYSE market, including Defe. & Aero-Retailer, Oil-Vehicle, Utility-IG and Healthcare-Machinery. Three cluster pairs are detected for the SSE market, i.e., IT-BM, Utility-Finance and IT-Trad(BC). The cluster pair, which can not be observed in the community structure of the full cross-correlation matrix, is a new finding in the financial systems.

To support the procedure of taking the absolute values of the negative elements of the sector-mode cross-correlation matrix C_{sec} , we set those negative elements to zeros or small positive values. The PMFG graph is then generated from the modified cross-correlation matrix of the sector mode, and the community structure is extracted. The resulting business sectors are similar to those in Fig. 5, except for the absence of the cluster pairs, due to the removing of the negative correlations. These results support the procedure of taking the absolute values of the negative matrix elements.

To quantify the local interactions of the business sectors in the sector mode, the average correlations inside a business sector and between two business sectors are computed according to Eqs. (1) and (2), for the cross-correlation matrices C_{sec} and $|C_{sec}|$, respectively. The matrix elements of $|C_{sec}|$ are just the absolute values of those of C_{sec} defined in Eq. (8). The results are shown in Table III. Since the market mode has been removed, the magnitude of the correlation becomes relatively small. However, the differences between various average correlations are very clear.

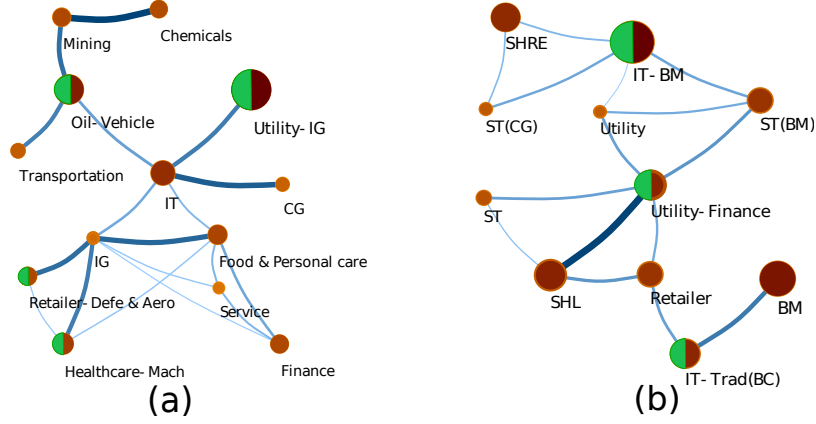


FIG. 5. The interaction structure of the business sectors in the sector mode, obtained from the PMFG graph. Two semi-circles in a full circle represent a cluster pair, and the label A-B denotes that the business sectors A and B form a cluster pair. (a) and (b) display the results for the NYSE and SSE markets respectively. The abbreviations are as follows. Deft & Aero: Defense and Aerospace, and Trad(BC): traditional industry of the Blue-chips stocks.

The average correlation inside the business sectors is stronger than that between the business sectors. Meanwhile, the average correlation between the business sectors with a link is stronger than that without a link. It should be noted that the correlation between the positive and negative subsectors in a cluster pair is negative, and this also results in that the average correlation inside the business sectors for $|C_{sec}|$ is larger than that for C_{sec} .

In particular, an important observation is that for the sector-mode cross-correlation matrix C_{sec} , the average correlation inside the business sectors is positive, while that between the business sectors is negative. Considering that the average correlation for C_{sec} is zero, the negative average correlation between the business sectors indicates that the business sectors are anti-correlated each other in the sector mode, and this is a somewhat surprising result. This kind of anti-correlations can not be directly detected in the full cross-correlation matrix, due to the dominating effect of the market mode. This should be potentially important in both theoretical and practical senses.

Time evolution of interaction structure

The time evolution of the interaction structure of the business sectors is an important topic in financial dynamics. Practically, it is also useful for investors to diversify risk in stock markets. Now, we investigate how the interaction structure evolves with time, combining the complex network methods with the random matrix decomposition. For the NYSE stock market, the time period of the daily data of the 259 stocks is equally divided into four time windows, i.e., from Jan., 1990 to Mar., 1994, from Apr., 1994 to Jun., 1998, from Jul., 1998 to Sep., 2002, and from Oct., 2002 to Dec., 2006, which are denoted by window (a), (b), (c) and (d), respectively. In the NYSE market, there is a typical financial bubble, i.e., the Dotcom bubble, which is mainly located in the window (c). The Dotcom bubble is also referred to as the Internet bubble or information technology bubble.

In Fig. 6, the interaction structure of the full cross-correlation matrix generated by the PMFG graph is shown for the four time windows. Clearly, the interaction structure for different time windows shares common features. For example, there are two large cores for all the four time windows, and most of other communities connect to the cores. In other words, the interaction structure of the business sectors is dominated by the cores. These cores are mainly induced by the market mode, almost time-independent.

For comparison, the local interaction structure of the business sectors in the sector mode is also extracted and displayed for the four time windows in Fig. 7. Obviously, it varies with time, in contrast to that obtained from the full cross-correlation matrix. This time-dependent interaction structure is in spirit consistent with the result in Ref. [43], in which the network structure of individual stocks is very different before, during and after a financial crisis. More specifically, the interaction structure of the window (c) is very simple, compared to others. The IT sector plays an important role in the window (c), while it is not so important in other windows. We believe that the simple interaction structure is driven by the Dotcom bubble in that period of time. This IT-dominating structure is reasonable, since the IT industry is very booming and becomes the core of the market during the Dotcom bubble.

A similar result is also obtained for the SSE stock market. We equally divide the time series of the daily data of the 259 stocks into two time windows. The window (a) is a standard period, from Jan., 1997 to Mar., 2002, and the window (b) is from Apr., 2002 to Nov., 2007, which contains a huge financial bubble from year 2003 to 2007. The

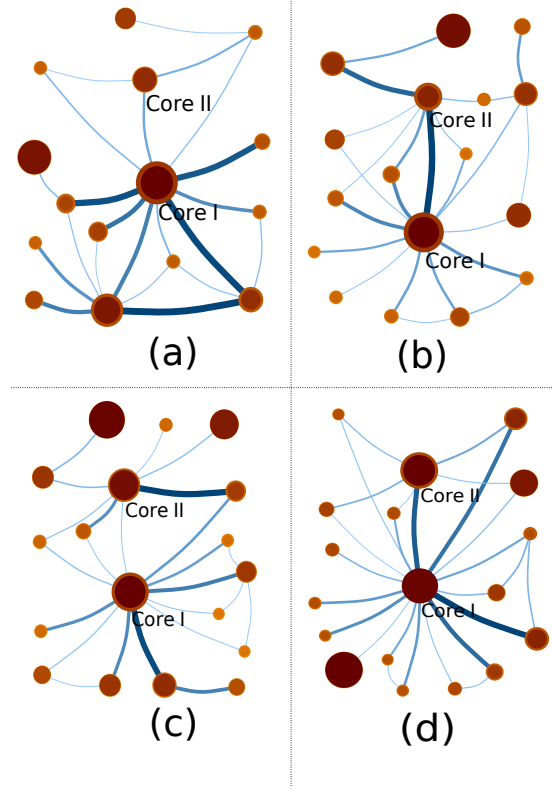


FIG. 6. The interaction structure of the business sectors for the full cross-correlation matrix is obtained with the PMFG graph for the NYSE market in different time windows. The business sectors in different time windows consist of similar stocks. Window (a): Jan., 1990 to Mar., 1994; window (b): Apr., 1994 to Jun., 1998; window (c): Jul., 1998 to Sep., 2002; window (d): Oct., 2002 to Dec., 2006. The Dotcom bubble is mainly located in the window (c).

local interaction structure of the business sectors in the sector mode extracted for the two time windows is displayed in Fig. 8. Clearly, the interaction structure of the window (b) is simple, and the SHRE (Shanghai Real Estate) sectors including the Basic Material sector are dominating. Actually, the huge bubble in the window (b) is mainly induced by the booming of the real estate industry during that time period.

To further support our results, we analyze another data set, which contains the daily data of the 249 stocks from Oct., 2007 to Nov., 2008, for both the NYSE and SSE markets. During this time period, both stock markets experienced the subprime mortgage crisis. The local interaction structure of the business sectors in the sector mode is shown in Fig. 9. A relatively simple structure is again observed for both stock markets. Since the subprime mortgage crisis is induced from finance and banking, the Finance sector is important in the interaction structure of the business sectors, especially in the NYSE market. Further, the crisis leads to an economic recession, and the ST sector becomes dominating for the SSE market.

Briefly speaking, the interaction structure of the business sectors in the sector mode changes dramatically in different time periods. Especially, it is rather simple during a financial bubble or crisis, dominated by specific business sectors, which are associated with the corresponding economic situation.

Finally, we investigate whether there exist interactions between the NYSE and SSE markets at the level of the business sectors. For this purpose, we combine the NYSE and SSE markets to form a single market. The overlapping time period of the data sets for the two markets is from Jan., 1997 to Dec. 2006. We repeat the analysis as described in the previous texts. In Fig. 10, the interaction structure of the business sectors is displayed. Although some stocks of the SSE market may mix into the business sectors of the NYSE market and vice versa, no interactions at the level of the business sectors are detected between the NYSE and SSE markets, for either the full cross-correlation matrix or the sector mode. The only link connecting the two markets is from the method itself, i.e., the graph should not be disconnected. However, one observes that the number of the business sectors in the sector mode is greatly reduced, and more cluster pairs emerge. This is because the cross-correlations between the NYSE and SSE markets are weak, and the stocks tend to form larger communities. We have also performed such an analysis in different time windows, including that with a financial bubble or crisis. The results remain qualitatively unchanged. In other words, the interactions between the NYSE and SSE markets are not so strong in the past years, at least in the level of the

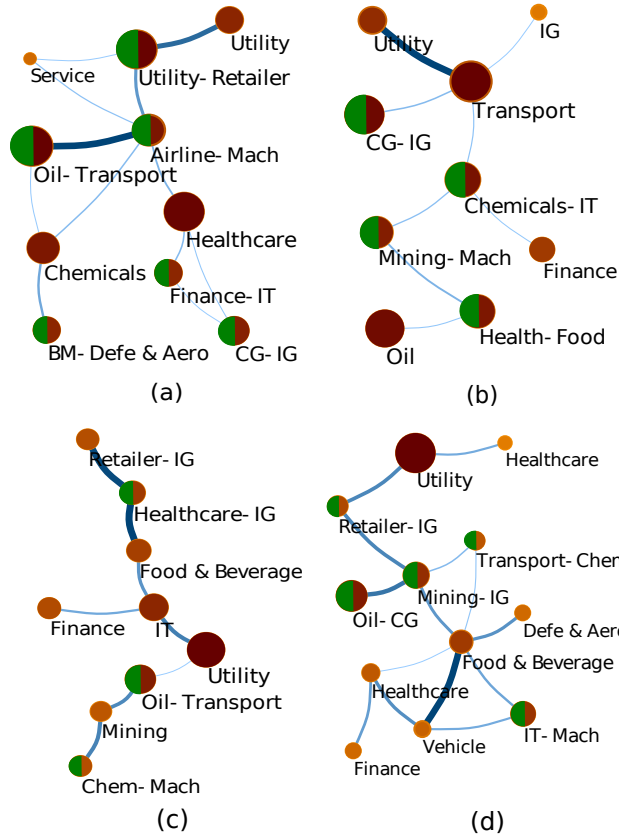


FIG. 7. The interaction structure of the business sectors in the sector mode is obtained with the PMFG graph for the NYSE market in different time windows. The time windows are the same as in Fig. 6.

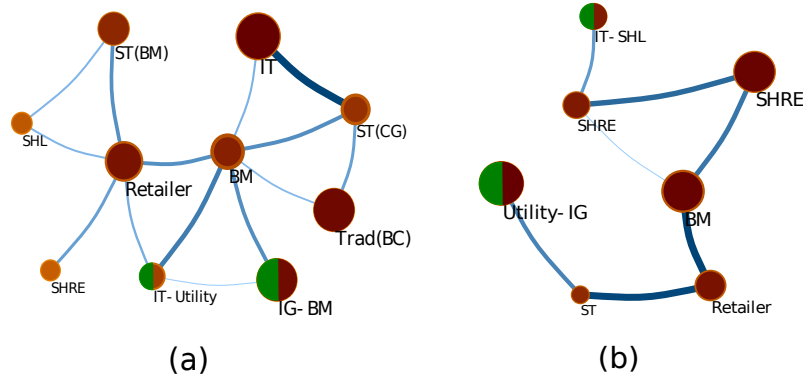


FIG. 8. The interaction structure of the business sectors in the sector mode is obtained with the PMFG graph for the SSE market in two time windows. Window (a): Jan., 1997 to Mar., 2002; window (b): Apr., 2002 to Nov., 2007. The huge bubble induced by the real estate industry is contained in the window (b).

business sectors.

DISCUSSION

In this paper, we investigate the communities and their interactions in complex financial systems, with the concepts and methods in statistical mechanics. We generate the interaction networks from the cross-correlation matrix of the financial market, based on the MST tree and PMFG graph. With the Infomap method, we extract the communities and their interaction structure. In the financial markets, the communities could be associated with the business

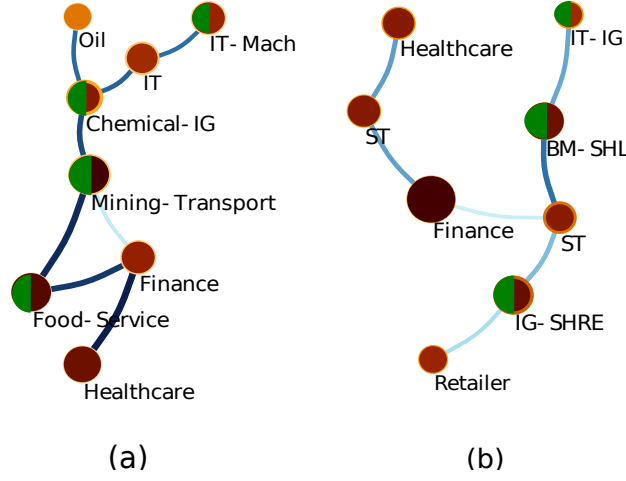


FIG. 9. The interaction structure of the business sectors in the sector mode is obtained with the PMFG graph in the time period of the subprime mortgage crisis. (a): for the NYSE market, from Oct., 2007 to Nov, 2008; (b): for the SSE market, from Oct., 2007 to Nov., 2008.

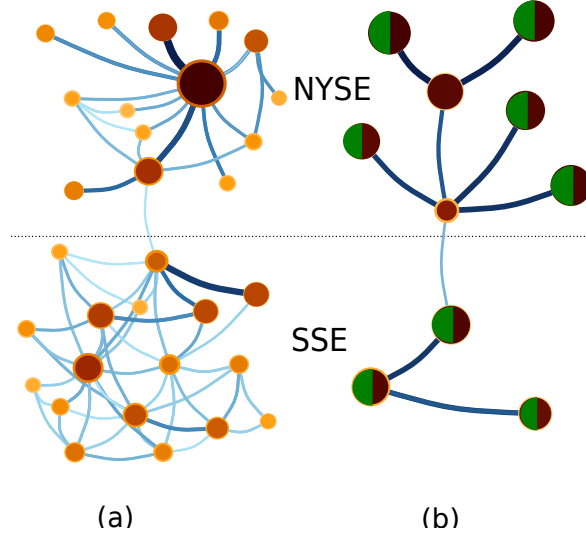


FIG. 10. The interaction structure of the business sectors is obtained with the PMFG graph for a combination of the NYSE and SSE markets. (a) and (b) are for the full cross-correlation matrix and the sector mode respectively. The business sectors of two markets are separated with the dashed line.

sectors (subsectors). Thus the economic implication of the interaction structure of the communities is revealed.

Based on the random matrix decomposition, the market mode and random mode of the price movement can be subtracted. Thus we demonstrate that the local interactions of the business sectors are mainly contained in the sector mode, from the perspective of both the domain structure and the PMFG graph. In the sector mode, the average correlation inside a business sector is positive, while that between two business sectors is negative. In other words, the business sectors are anti-correlated each other. Meanwhile, a cluster pair structure is also observed. The interaction structure of the business sectors in the sector mode varies dramatically with time. In particular, it becomes simple during a financial bubble or crisis, dominated by specific business sectors. The interaction structure of the business sectors of the full cross-correlation matrix does not change so much with time.

METHODS

Random matrix decomposition

The logarithmic price return of the i -th stock over a time interval Δt is defined by

$$R_i(t) \equiv \ln P_i(t + \Delta t) - \ln P_i(t), \quad (3)$$

where $P_i(t)$ represents the stock price at time t . For the daily data, we typically take $\Delta t = 1$. To ensure different stocks with an equal weight, we introduce the normalized price return

$$r_i(t) = \frac{R_i(t) - \langle R_i(t) \rangle}{\sigma_i}, \quad (4)$$

where $\langle \cdots \rangle$ is the average over time t , and $\sigma_i = \sqrt{\langle R_i^2 \rangle - \langle R_i \rangle^2}$ denotes the standard deviation of R_i . Then, the matrix elements of the cross-correlation matrix C are defined by the equal-time correlations

$$C_{ij} \equiv \langle r_i(t) r_j(t) \rangle. \quad (5)$$

By the definition, C is a real symmetric matrix with $C_{ii} = 1$, and C_{ij} is valued in the range $[-1, 1]$.

Statistical properties of the eigenvalues of the so-called Wishart matrix are known, and the Wishart matrix is derived from uncorrelated time series with a finite length. We denote the total number of stocks by N , and the total time length of the data by T . In the limit $N \rightarrow \infty$ and $T \rightarrow \infty$ with $Q \equiv T/N \geq 1$, the probability distribution of the eigenvalues λ is given by [57, 58]

$$P_{rm}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_{max}^{ran} - \lambda)(\lambda - \lambda_{min}^{ran})}}{\lambda}, \quad (6)$$

with the upper and lower bounds $\lambda_{min(max)}^{ran} = [1 \pm (1/\sqrt{Q})]^2$.

Large eigenvalues of the cross-correlation matrix in financial markets deviate from $P_{rm}(\lambda)$ of the Wishart matrix. It suggests that there exist non-random interactions between individual stocks, and the largest eigenvalue represents the global price movement of the market. In our notation, the eigenvalues are arranged in the order of $\lambda_\alpha > \lambda_{\alpha+1}$, with $\alpha = 0, \dots, N$. The eigenmodes with large eigenvalues can be associated with business sectors. Taking into account the signs of the components of the eigenmodes, the sector of λ_α may be separated into two subsectors, i.e., the positive and negative subsectors [16].

Finally, based on the RMT theory, the cross-correlations C_{ij} can be decomposed into different eigenmodes,

$$C_{ij} = \sum_{\alpha=1}^N \lambda_\alpha C_{ij}^\alpha, \quad C_{ij}^\alpha = u_i^\alpha u_j^\alpha, \quad (7)$$

where λ_α is the α -th eigenvalue of C_{ij} , u_i^α is the i -th component in the α -th eigenvector. Thus, C_{ij}^α represents the cross-correlation in the α -th eigenmode. In order to uncover the interaction structure in different eigenmodes, we define three mode cross-correlation matrices by

$$C_{mode,ij} = \sum_{\alpha}^{mode} C_{ij}^\alpha. \quad (8)$$

For the market mode, $C_{mar,ij} = C_{ij}^0$, corresponding to the largest eigenvalue of C_{ij} . For the sector mode, $C_{sec,ij} = \sum_{\alpha=1}^{11} C_{ij}^\alpha$ for the NYSE market, and $C_{sec,ij} = \sum_{\alpha=1}^6 C_{ij}^\alpha$ for the SSE market. For the random mode, we simply take $C_{ran,ij} = \sum_{\alpha=50}^{258} C_{ij}^\alpha$ for both the NYSE and SSE markets.

Methods of complex networks

In principle, the matrix elements C_{ij} contain full information of the cross-correlations between stocks, but carry also strong noises. To sketch the backbone of the interactions between stocks, the edges should be greatly discarded with respect to the complete graph. To this aim, the graph should be kept as simple as possible, but important

correlations should remain. The MST tree is therefore the simplest connected graph [59, 60], in which there are no cycles. Moreover, the MST tree may give the simplest subdominant organization structure of individual stocks in a direct scheme.

We briefly describe the MST method as follows: (1) Rank the correlations C_{ij} between stocks i and j in a list by descending order. (2) Capture the first one in the list, and add the nodes i and j , and the edge between i and j to the graph. (3) Take the next one in the list, and add the corresponding edges and nodes to the graph if it could be linked to the existing nodes, and would not form cycles; otherwise ignore it. (4) Repeat step (3) until all links have been visited.

Assuming that there are N stocks, the resulting MST tree consists of N nodes and $N - 1$ edges, and maximizes the sum of C_{ij} in the tree graph. The MST tree is an efficient approach to capture the most relevant correlations of each node. However, it is a strong constraint that only the tree structure is permitted. All those edges are rejected when they form cycles with the existing edges, even if they may represent significant correlations.

To improve the method, the PMFG graph is proposed [42]. The solution generates a graph embedded on a surface with a particular genus g . The genus of a surface is the number of holes in the surface, and $g = 0$ corresponds to a topological sphere, $g = 1$ to a torus, $g = 2$ to a double torus, etc. The PMFG algorithm is generally the same as the MST one, except for the step (3). It is replaced by the condition that the resulting graph should be embeddable on a surface with genus g . The graph generated by the PMFG algorithm is a triangulation of the surface, and it contains $3N + 6g - 6$ links, which maximize the sum of C_{ij} . The simplest case is the graph with $g = 0$. In some sense, the PMFG graph could be viewed as the minimum extension towards the complexity after the MST tree.

The next step after the extraction of the MST tree and PMFG graph, we apply the Infomap method proposed in Ref. [47], to investigate the community structure of the financial network. The method is based on a random walk on the network to capture the information flow, then to extract the community structure from the regularity of the data. Detailed descriptions of the method are referred to Ref. [47]. With the Infomap method, we can identify the community structure for both the MST tree and PMFG graph.

Further, with the mode cross-correlation matrix C_{mode} at hand, we may construct the MST tree and PMFG graph for different modes, and investigate which modes dominate the business sectors and their interactions. This is an important step in the exploration of the interaction structure in financial systems.

ACKNOWLEDGMENTS

This work was supported in part by NNSF of China under Grant Nos. 11375149 and 11075137, and Zhejiang Provincial Natural Science Foundation of China under Grant No. Z6090130.

-
- [1] Mantegna, R.N., & Stanley, H.E., Scaling behavior in the dynamics of an economic index, *Nature* **376**, 46-49 (1995).
 - [2] Gopikrishnan, P., Plerou, V., Amaral, L.A.N., Meyer, M. & Stanley, H.E., Scaling of the distribution of fluctuations of financial market indices, *Phys. Rev. E* **60**, 5305 (1999).
 - [3] Giardina, I., Bouchaud, J.P., & Mézard, M., Microscopic models for long ranged volatility correlations, *Physica A* **299**, 28-39 (2001).
 - [4] Bouchaud, J.P., Matalcz, A., & Potters, M., Leverage effect in financial markets: The retarded volatility model, *Phys. Rev. Lett.* **87**, 228701 (2001).
 - [5] Bouchaud, J.P. & Potters, M., *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, Cambridge University Press, England, 2003.
 - [6] Sornette, D., Critical market crashes, *Phys. Rep.* **378**, 1-98 (2003).
 - [7] Gabaix, X., Gopikrishnan, P., Plerou, V. & Stanley, H.E., A theory of power-law distributions in financial market fluctuations, *Nature* **423**, 267-270 (2003).
 - [8] Qiu, T., Zheng, B., Ren, F. & Trimper, S., Return-volatility correlation in financial dynamics, *Phys. Rev. E* **73**, 065103 (2006).
 - [9] Garas, A., Argyrakis, P. & Havlin, S., The structural role of weak and strong links in a financial market network, *Eur. Phys. J. B* **63**, 265-271 (2008).
 - [10] Shen, J. & Zheng, B., Cross-correlation in financial dynamics, *Europhys. Lett.* **86**, 48005 (2009).
 - [11] Shen, J. & Zheng, B., On return-volatility correlation in financial dynamics, *Europhys. Lett.* **88**, 28003 (2009).
 - [12] Podobnik, B., Horvatić, D., Petersen, A.M. & Stanley, H.E., Cross-correlations between volume change and price change, *Proc. Natl. Acad. Sci.* **106**, 22079-22084 (2009).
 - [13] Podobnik, B., Wang, D., Horvatic, D., Grosse, I. & Stanley, H.E., Time-lag cross-correlations in collective phenomena, *EPL* **90**, 68001 (2010).

- [14] Podobnik, B., Valentinčič, A., Horvatić, D. & Stanley, H.E., Asymmetric lévy flight in financial ratios, *Proc. Natl. Acad. Sci.* **108**, 17883-17888 (2011).
- [15] Li, W., Wang, F.Z., Havlin, S. & Stanley, H.E., Financial factor influence on scaling and memory of trading volume in stock market, *Phys. Rev. E* **84**, 046112 (2011).
- [16] Jiang, X.F., & Zheng, B., Anti-correlation and subsector structure in financial systems, *EPL* **97**, 48006 (2012).
- [17] Kumar, S. & Deo, N., Correlation and network analysis of global financial indices, *Phys. Rev. E* **86**, 026101 (2012).
- [18] Preis, T., Kenett, D.Y., Stanley, H.E., Helbing, D. & Ben-Jacob, E., Quantifying the behavior of stock correlations under market stress, *Sci. Rep.* **2**, 752 (2012).
- [19] Jiang, X.F., Chen, T.T. & Zheng, B., Time-reversal asymmetry in financial systems, *Physica A* **392**, 5369-5375 (2013).
- [20] Münnix, M.C. et al., Identifying states of a financial market, *Sci. Rep.* **2**, 644 (2012).
- [21] Górski, A.Z., Drożdż, S. & Speth, J., Financial multifractality and its subtleties: an example of DAX, *Physica A* **316**, 496-510 (2002).
- [22] Drożdż, S., Grümmner, F., Górski, A.Z., Ruf, F. & Speth, J., Dynamics of competition between collectivity and noise in the stock market, *Physica A* **287**, 440-449 (2000).
- [23] Chen, J.J., Zheng, B. & Tan, T., Agent-based model with asymmetric trading and herding for complex financial systems, *PLoS One* **8**, e79531 (2013).
- [24] Ouyang, F.Y., Zheng, B. & Jiang, X.F., Spatial and temporal structures of four financial markets in greater china, *Physica A* **402**, 236-244 (2014).
- [25] Bollen, J., Mao, H. & Zeng, X. J. , Twitter mood predicts the stock market, *J. of Comp. Sci.* **2**, 1-8 (2011).
- [26] Preis, T., Moat, H. S. & Stanley, H. E., Quantifying Trading Behavior in Financial Markets Using Google Trends, *Sci. Rep.* **3**, 1684 (2013).
- [27] Moat, H. S. et al., Quantifying Wikipedia Usage Patterns Before Stock Market Moves, *Sci. Rep.* **3**, 1801 (2013).
- [28] Mantegna, R.N. & Stanley, H.E., Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, England, 2000.
- [29] Erb, C. B., Harvey, C. R. & Viscanta, T. E., Forecasting international equity correlations, *Financ. Anal. J.* **50**, 32-45 (1994).
- [30] Solnik, B., Bourcuelle, C. & Fur, Y. Le, International market correlation and volatility, *Financ. Anal. J.* **52**, 17-34 (1996).
- [31] LeBaron, B., Arthur, W. B. & Palmer, R., Time series properties of an artificial stock market, *J. Econ. Dyn. Control* **23**, 1487-1516 (1999).
- [32] Laloux, L., Cizeau, P., Bouchaud, J.P. & Potters, M., Noise Dressing of Financial Correlation Matrices, *Phys. Rev. Lett.* **83**, 1467 (1999).
- [33] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N. & Stanley, H.E., Universal and nonuniversal properties of cross correlations in financial time series, *Phys. Rev. Lett.* **83**, 1471 (1999).
- [34] Plerou, V. et al., Random matrix approach to cross correlations in financial data, *Phys. Rev. E* **65**, 066126 (2002).
- [35] Utsugi, A., Ino, K. & Oshikawa, M., Random matrix theory analysis of cross correlations in financial markets, *Phys. Rev. E* **70**, 026110 (2004).
- [36] Qiu, T., Zheng, B. & Chen, G, Financial networks with static and dynamic thresholds, *New J. Phys.* **12**, 043057 (2010).
- [37] Oh, G. et al., Statistical properties of cross-correlation in the korean stock market, *Euro. Phys. J. B* **79**, 55-60 (2011).
- [38] Pan., R.K. & Sinha, S., Collective behavior of stock price movements in an emerging market, *Phys. Rev. E* **76**, 046116 (2007).
- [39] Mantegna, R.N., Hierarchical structure in financial markets, *Euro. Phys. J. B* **11**, 193-197 (1999).
- [40] Schweitzer, F. et al., Economic networks: The new challenges, *Science* **325**, 422-425 (2009).
- [41] Kenett, D.Y., Preis, T., Gur-Gershgoren, G. & Ben-Jacob, E., Dependency network and node influence: Application to the study of financial markets, *Int. J. Bifur. Chaos* **22**, 1250181 (2012).
- [42] Tumminello, M., Aste, T., Di Matteo, T. & Mantegna, R.N., A tool for filtering information in complex systems, *Proc. Natl. Acad. Sci. USA* **102**, 10421-10426 (2005).
- [43] Aste, T., Shaw, W. & Di Matteo T., Correlation structure and dynamics in volatile markets, *New J. Phys.* **12**, 085009 (2010).
- [44] Pozzi, F., Di Matteo, T. & Aste, T., Centrality and peripherality in filtered graphs from dynamical financial correlations, *Adv. Complex Syst.* **11**, 927 (2008).
- [45] Di Matteo, T., Pozzi, F. & Aste, T., The use of dynamical networks to detect the hierarchical organization of financial market sectors, *Euro. Phys. J. B* **73**, 3-11 (2010).
- [46] Kenett, D.Y. et al., Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market, *PLoS ONE* **5**, e15032 (2010).
- [47] Rosvall, M. & Bergstrom, C.T., Maps of random walks on complex networks reveal community structure, *Proc. Natl. Acad. Sci.* **105**, 1118-1123 (2008).
- [48] Newman, M.E.J., & Girvan, M., Finding and evaluating community structure in networks, *Phys. Rev. E* **69**, 026113 (2004).
- [49] Newman, M.E.J., Modularity and community structure in networks, *Proc. Natl. Acad. Sci.* **103**, 8577-8582 (2006).
- [50] Li, H.J. et al., Community structure detection based on potts model and network's spectral characterization, *EPL* **97**, 48005 (2012).
- [51] Galbiati, M., Delpini, D. & Battiston, S. , The power to control, *Nat. Phys.* **9**, 126-128 (2013).
- [52] Nicosia, V., Criado, R., Romance, M., Russo, G. & Latora G., Controlling centrality in complex networks, *Sci. Rep.* **2**, 218 (2012).
- [53] Battiston, S., Puliga, M., Kaushik, R., Tasca, P. & Caldarelli, G., Debtrank: Too central to fail? financial networks, the

TABLE I. The PMFG communities and the corresponding RMT subsectors. λ_α^+ and λ_α^- represent the positive and negative subsectors, respectively, in the α -th eigenmode of the RMT theory [16]. The abbreviations are as follows. Info. service: Information service, Japan-IT: Japan IT company, BM: Basic materials, CG: Consumer goods, mach.: machinery, SHRE: Shanghai Real Estate, SHL: Shanghai Local company, IG: Industrial goods, ST (BM&IG): BM and IG with ST, ST (FI&CG): Finance and CG with ST.

NYSE		SSE	
PMFG	RMT	PMFG	RMT
Utility	λ_1^+	Finance	λ_5^-
Transportation	λ_{11}^-	IT	λ_2^-
Oil	λ_2^-	SHRE	λ_3^-
Gold mining	λ_5^-	ST (BM&IG)	λ_1^-
Retailer	λ_7^-, λ_8^+	ST (FI&CG)	λ_3^-
Info. service	λ_1^-	IG	λ_6^+
Technology	λ_3^+, λ_7^+	CG	
Japan-IT	λ_{10}^-	Cement	
Healthcare	λ_8^-	BM	
CG	λ_3^-	SHL	
Insurance	λ_6^-, λ_9^+	Utility	
Bank	λ_6^-, λ_9^+	Retailer	
Airline	λ_{11}^-	Healthcare	
BM			
Diversity mach.			
Construction mach.			
Electrical equipment			
Defense&Aerospace			
Vehicle			
Entertainment			

TABLE II. The average cross-correlation \overline{C}_{ij} for the PMFG graph, MST tree and RMT sectors (subsectors). \overline{C}_{ij}^{in} and \overline{C}_{ij}^{be} represent the average correlations inside and between the communities respectively. \overline{C}_{ij}^{li} and \overline{C}_{ij}^{de} are the average correlations between two different communities with a link and without a link. \overline{C}_{ij}^{+-} is the average correlation between the positive and negative subsectors in an eigenmode [16].

	\overline{C}_{ij}	\overline{C}_{ij}^{in}	\overline{C}_{ij}^{be}	\overline{C}_{ij}^{li}	\overline{C}_{ij}^{de}	\overline{C}_{ij}^{+-}
NYSE PMFG	0.16	0.46	0.15	0.18	0.13	
MST	0.16	0.49	0.16	0.19	0.12	
RMT	0.16	0.33	0.16			0.15
SSE PMFG	0.37	0.43	0.36	0.38	0.32	
MST	0.37	0.48	0.36	0.39	0.31	
RMT	0.37	0.40	0.34			0.32

fed and systemic risk, *Sci. Rep.* **2**, 541 (2012).

- [54] Nepusz, T. & Vicsek, T., Controlling edge dynamics in complex networks, *Nat. Phys.* **8**, 568-573 (2012).
- [55] Vitali, S., Glattfelder, J. & Battiston, S., The network of global corporate control, *PloS one* **6**, e25995 (2011).
- [56] Wilcox, D. & Gebbie, T., An analysis of cross-correlations in an emerging market, *Physica A* **375**, 584-598 (2007).
- [57] Dyson, F.J., Distribution of eigenvalues for a class of real symmetric matrices, *Rev. Mex. Fis.* **20**, 231-237 (1971).
- [58] Sengupta, A.M. & Mitra, P.P., Distributions of singular values for some random matrices, *Phys. Rev. E* **60**, 3389 (1999).
- [59] Kruskal, J.B., On the shortest spanning subtree of a graph and a traveling salesman problem, *Proc. Am. Math. Soc.* **7**, 48-50 (1956).
- [60] Prim, R.C., Shortest connection networks and some generalizations, *Bell Syst. Tech. J.* **36**, 1389-1401 (1957).

AUTHOR CONTRIBUTION

X.F.J. and B.Z. conceived the study. X.F.J., T.T.C and B.Z. designed and performed the research. X.F.J. and T.T.C performed the statistical analysis of the data. X.F.J. drafted the manuscript. B.Z. reviewed and approved the manuscript.

TABLE III. The average cross-correlation \overline{C}_{ij} of the sector mode for the PMFG graph. All values in the table have been multiplied by a factor of 10^2 . \overline{C}_{ij}^{in} and \overline{C}_{ij}^{be} represent the average correlations inside and between the communities respectively. \overline{C}_{ij}^{li} and \overline{C}_{ij}^{de} are the average correlations between two different communities with a link and without a link. \overline{C}_{ij}^{+-} denotes the average correlation between the positive and negative subsectors in a cluster pair.

		\overline{C}_{ij}	\overline{C}_{ij}^{in}	\overline{C}_{ij}^{be}	\overline{C}_{ij}^{li}	\overline{C}_{ij}^{de}	\overline{C}_{ij}^{+-}
NYSE	C_{sec}	0	7.4	-0.7	-0.4	-0.8	-0.9
	$ C_{sec} $	3.2	9.7	2.5	2.8	2.4	2.3
SSE	C_{sec}	0	1.3	-0.2	-0.1	-0.2	-0.2
	$ C_{sec} $	2.1	3.8	1.9	2.3	1.8	1.7

ADDITIONAL INFORMATION

Competing financial interests: The authors declare no competing financial interests.